Flash Express
A Performance Snapshot

IBM

Introduction

As companies compete more aggressively in the new economy, service attributes create a base upon which companies can differentiate themselves. Most companies require high availability and performance of their key workloads and cannot tolerate variability in service levels. Flash memory on the IBM zEnterprise® EC12 (zEC12) server is used as a means of significantly improving the availability and consistent performance of your key systems that cannot afford availability or performance disruptions. For customers that are measured on their service levels and especially those who might face penalties if SLAs are unmet, having Flash can boost availability and performance levels.

With Flash Express, critical workload transitions can occur faster, allowing transactions to reach normal steady state rapidly with resulting improved transactional response times. In addition, fast page-ins from Flash Express, means that diagnostic capture time, such as with SVC dumps, can be dramatically reduced with the performance impact on other workloads minimized.

Java™ CPU performance can be enhanced, as well, using pageable large pages which are available only on systems using Flash Express. Flash Express represents a way to improve availability and performance with a simple solution that is readily configurable, easy to use, and offers a fast return on investment.

What is Flash Express?

Flash consists of non-volatile storage using solid state devices on a PCIe card form factor. Flash SSDs have no moving mechanical parts, no rotational delays and Flash technology is designed to offer latencies that are far superior to the fastest hard disk drives. As such, it offers superior performance characteristics as compared to spinning disk.

Flash Express is an optional feature introduced with the zEC12, to drive performance improvements and better availability for critical business workloads that cannot afford variability in service levels.

Flash Express implements a new tier of memory, called Storage Class Memory (SCM), made accessible through Flash Express.

Flash Cards are available (separately priced feature) in card pairs providing mirrored data to ensure high availability and redundancy. A maximum of four pairs of cards can be installed on a single zEC12, for a maximum capacity of 1.4 TB usable storage per card pair, or 5.6 TB of storage for 4 card pairs in total.

Flash Express, unlike other applications of Flash memory, is used server side to aid in performance of paging. When used with pageable large pages there is an additional performance benefit realized. Managing 1MB vs. 4KB memory “chunks” provides application performance benefits by reducing memory management costs, especially for today’s data intensive applications.

Security

Companies need to protect their sensitive data when it resides on persistent storage such as Flash Express. The zEC12 server automatically encrypts all data stored in Flash Express memory. Hardware encryption can improve data security ensured through the use of a Smart Card and a unique key that is stored on the Support Element. The paging data on Flash Express is protected with the Smart Card key and can **only** be used on the system with the key that encrypted it. If the Smart Card is not presented, the data cannot be accessed, providing an additional layer of protection against unauthorized use. The use of customer managed smart cards can also help simplify end of life audit requirement and reduce management costs.

Benefits of Flash Express

z/OS® is the first exploiter to use Flash Express storage as Storage Class Memory (SCM) for paging. The z/OS paging subsystem works with a mix of internal Flash Express storage and External Disk. z/OS detects whether Flash Express storage is assigned to the LPAR and will try to page to Flash Express before using paging datasets on DASD and z/OS automatically uses Flash Express storage for paging unless otherwise specified.

Note that Flash Express is not intended to be a replacement for real memory. IBM recommends that customers purchase enough real memory to avoid most paging while running production work.

Flash Express can help performance when page frames must be paged in quickly. Though modern DASD devices are very fast, pages can be retrieved from Flash Express even faster. For example, when a workload transition occurs causing an inactive application to become active again quickly, this application may resume slowly due to paging. Or, an SVC dump may be delayed waiting for pages to be retrieved from DASD so they can be collected for the creation of a dump. In both cases, faster page-ins with Flash Express can significantly reduce latency delays.

Flash Express offers another benefit for system software exploiters as it is designed to allow the use of one MB *pageable* large pages. Pageable large pages can provide the same fast memory access as fixed large pages without requiring "pinning" of real memory and without requiring special security configuration.

Another exploiter for Flash Express is Linux on System z; the support has been made available to the Linux distributors. Flash Express is a vehicle to improve performance when writing files to temporary storage such as dump files. It is expected that this use of Flash Express will deliver accompanying performance benefits over writing to disk in Linux on System z environments. z/VM does not support Flash Express.

Tests

The following IBM tests compare the performance characteristics of Flash and Disk.

The first test represents a workload transition. On your system, this might be analogous to a transition from overnight batch work to Online Transaction Processing. The test used a WebSphere® Application Server (WAS) workload running CICS® transactions accessing data in DB2®. Initially three WAS instances were run with each WAS instance requiring 6.5 gigabytes of Java heap. After simulating a client load to three WAS instances, all activity to the WAS applications was stopped, and other work was run with the intent of driving WAS memory out to DASD or to Flash SCM. All performance information was determined in a controlled environment.

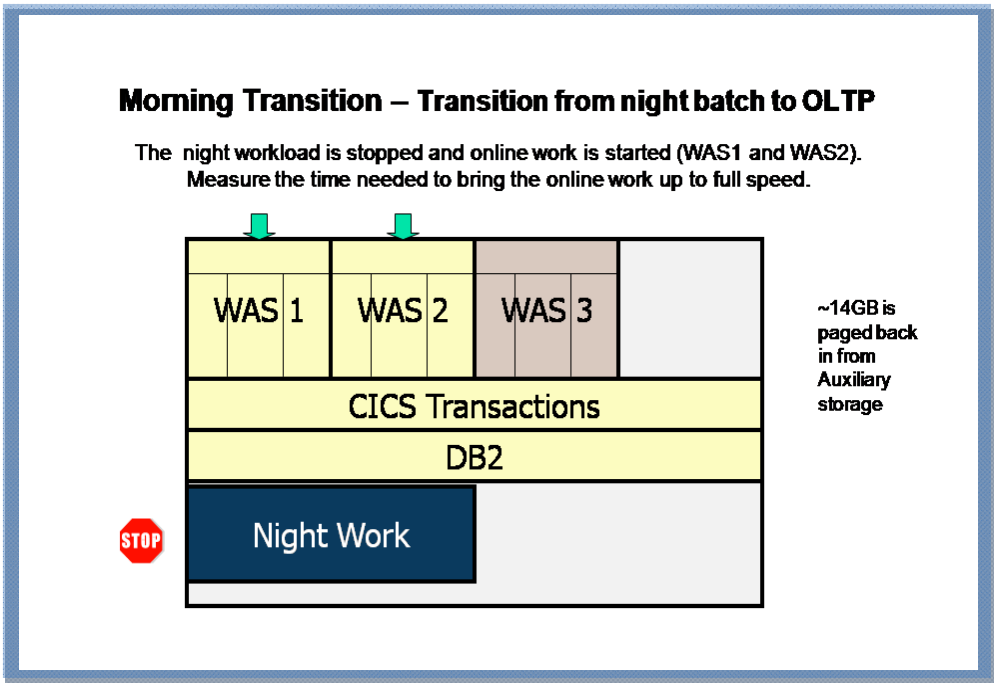

Figure 1. Configuration for tests

After some time, the WAS activity was resumed on two of these three WAS instances and measurements were taken to determine the time needed for the WAS workloads to reach peak transaction rates.

This workload transition test was run initially with paging to a dedicated DS8800 Model 2107-951 DASD device. Then the same test was run with paging to Flash SCM. Results were compared second by second to determine the differences in the workload transition time.

Observations

Running this workload transition scenario with paging to DASD, required about 44 seconds for the WAS instances to reach peak transaction rate. Running the same scenario with paging to Flash Express required about 10 seconds for the WAS instances to reach peak performance.

NOTE: When paging to Flash Express and SCM rather than to DASD, 37% more transactions were completed during the first 45 seconds with a 10 fold reduction in response time.

These performance boosts can be critical for your critical workloads and applications where heavy traffic predominates in the morning (as in service decks, trading applications or banking) and SLAs must be met in order to meet customer satisfaction levels.

Transaction ramp up

The chart below shows the transaction ramp up time for this morning transition scenario. The test results were captured second by second, with elapsed time depicted on the x-axis and transaction rate on the y-axis. When using paging to DASD, a gradual increase in transaction rate over 44 seconds before steady state transaction rate is reached. When paging to Flash Express, however, transaction rate ramped up much more quickly, in about 10 seconds, and with less variability. Workloads reached steady state in a quarter of the time they needed when using DASD.

Morning Transition Time Transaction Rate over Time

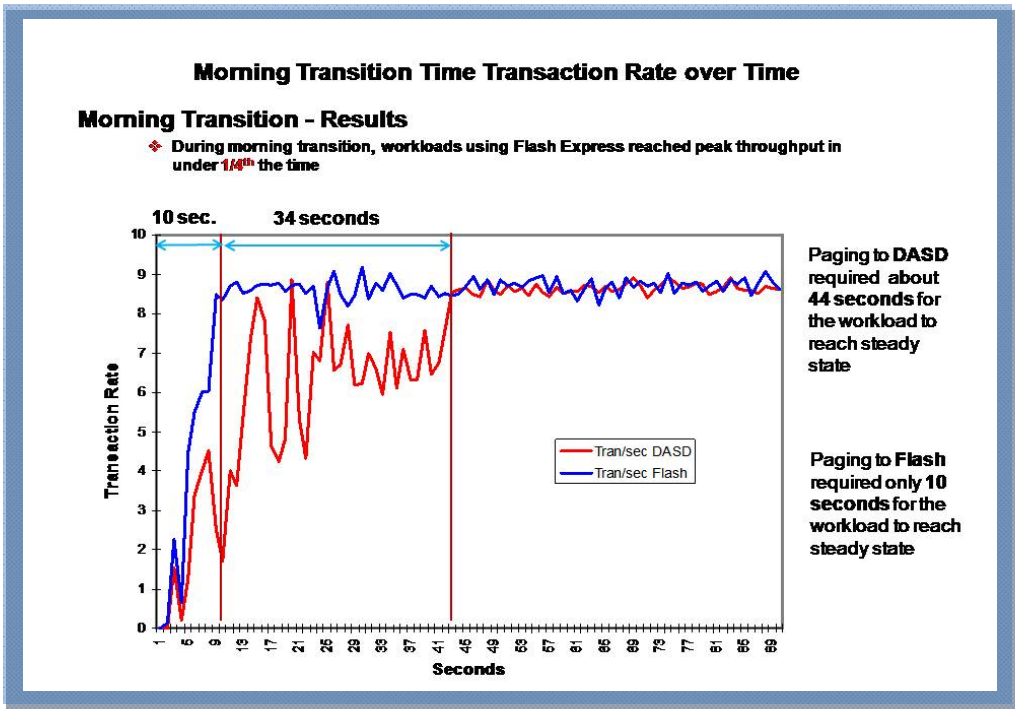

Figure 2. Morning Transition First 90 Seconds

- Notice that with DASD, the transactions do not achieve steady state until approximately 44 seconds; with Flash transactions achieve steady state at approximately 10 seconds, over a half minute earlier, or a 77% improvement.

- With Flash there is also a 90% reduction in response time and a 37% increase in throughput in the first 45 seconds of morning transition

SVC Dump Performance

In this second scenario an SVC (supervisor call) dump is executed. SVC dumps are needed to collect diagnostics, but work on the system can be delayed while the dump data is being captured. This is because pages needed for the dump must be collected, in their current state. If pages are on auxiliary storage when the dump is collected, these must be brought back into memory before being written to the dump. The time needed to bring these pages from auxiliary storage into memory can add seconds, and in some cases minutes, to dump capture time. This could significantly reduce the availability of your system and impact your SLAs.

- With Flash Express, pages can be paged into memory quickly, reducing dump capture time. Dumps completed almost 25% faster with Flash Express than they did when using DASD

To demonstrate this potential Flash Express benefit, the system was set up with four WAS instances, CICS and DB2. Each WAS instance had a control region and 3 servant regions with a total Java Heap requirement of 6.5 GB. Three of these WAS instances were active, but the 4th WAS instance was inactive, with most memory paged out. As the workload was running, an SVC dump was captured including 1 active WAS instance, 1 inactive WAS instance, and DB2.

With the SVC dump scenario, paging to Flash Express SCM reduced total dump capture time from roughly **60 seconds to about 14 seconds.** Much of the dump capture time was spent bringing pages into main memory from storage. By using Flash SCM for paging, page-in time was reduced which significantly cut total dump capture time by an impressive 76%.

Comparison of SVC Dump –DASD versus Flash Express

SVC Dump Metrics	DASD	Flash
SVC Dump size (in bytes):	18GB	18GB
DUMP Elapsed time:	189	143
Max address space non-dispatchable seconds	58.89	13.74
System non-dispatchable seconds	1.34	0.55

Figure 3. SVC Dump comparison statistics

- Using Flash Express in the SVC dump described above, Address Spaces were stopped 77% less time and Systems were stopped 59% less time as compared to when using DASD. This translates to higher availability when using Flash.

Diagnostic Collection

Most customers collect diagnostics such as SVC dumps which can be particularly disruptive when the necessary pages must be paged in to create a dump. During some phases of SVC dump processing, work is halted to allow this data to be collected before it changes. One phase stops work for the entire system, and this time is called *system-wide nondispatchability time*. Before this phase ends, the address spaces involved in dump data collection are stopped. Once work resumes for the rest of the system, those address spaces remain stopped until dump data collection ends. This time is called *address space nondispatchability*. The dump itself is written asynchronously in the background and is not impactful to performance.

- ***Flash Express can be used to significantly reduce both system-wide and address space nondispatchability times, improving availability for your system.***

Address Space Availability Differences during SVC Dump Using Flash and DASD

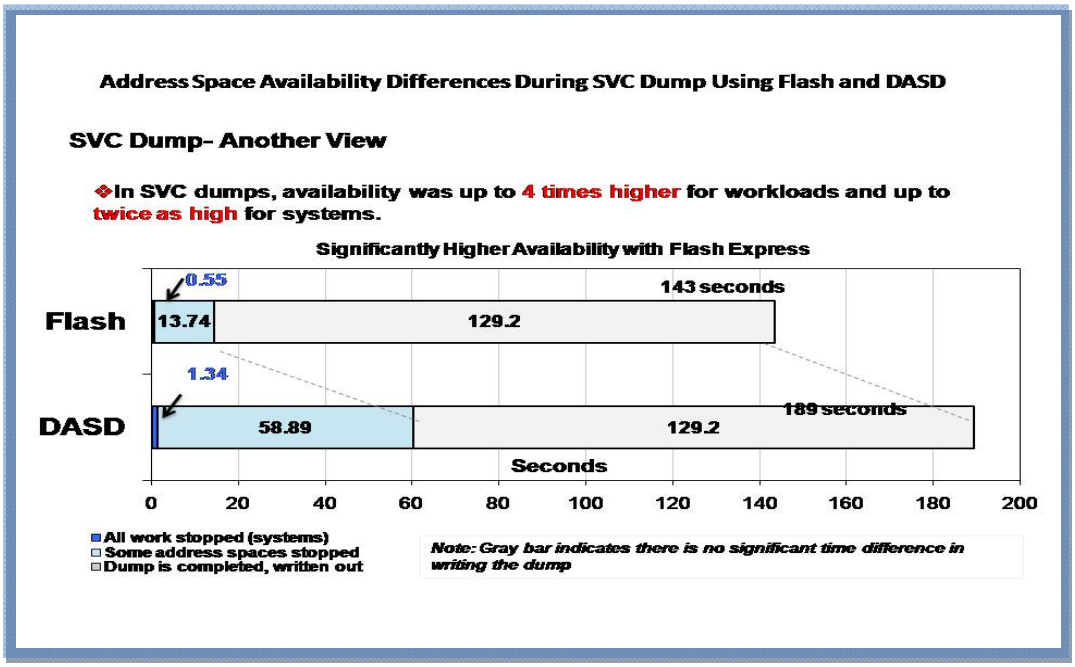

Figure 4. SVC Dump comparison chart showing non dispatchable time

Note: In our test scenario, some address spaces used about 2 GB of data on auxiliary storage. These address spaces were stopped until all the data on auxiliary storage could be brought back into memory. Flash is especially beneficial when DB2 or other critical address spaces are dumped. If these address spaces have data on auxiliary storage, paging to Flash can significantly reduce the time that critical address spaces are stopped.

- ***In internal IBM tests, using Flash Express showed reductions of over 59% in system-wide nondispatchability and over a 77% reduction in total address space nondispatchability time.***

Disruptions to availability can be reduced dramatically when Flash Express is configured.

When capturing an SVC dump, the time needed to bring data back in from auxiliary storage is critical. Work on your system may not be dispatched until this data can be paged back into memory. With Flash Express, data on auxiliary storage is paged in faster, allowing work to resume on your system sooner. The end result is higher system availability.

Standalone Dump Performance

In the case of standalone dump, the operating system must go through IPL in standalone dump mode to capture critical statistics. Usually this type of dump is used in more severe problem situations to capture an entire system image. A standalone dump involves copying all the contents of main memory to a storage device to record the information at the time of the problem for later problem debugging purposes. While SVC Dumps are collected, some other work is typically able to run on the system. But the system is completely stopped while a standalone dump is collected. The faster you can capture a standalone dump, the faster you can re-IPL your system and start processing work again.

As we can see the standalone dump benefited from paging to Flash Express. Note the read rate differences in the table below. There were some variations in the runs but Flash Express consistently brought pages in from auxiliary storage faster, significantly reducing the standalone dump time.

- ***Faster page-in with Flash reduced our standalone dump time by approximately 19%***

Standalone Dump tests

Tests	Total dump time in minutes	Paging I/O wait time in seconds	Read rate in MB/second	Total GB dumped	GB data read in from aux storage
DASD page data sets using DS8800	**00:03:12.92**	**00:00:41.30**	**438.06**	**36.2**	**17.7**
Flash Express used for paging	**00:02:35.03**	**00:00:10.38**	**1612.30**	**36.3**	**16.3**

Figure 5. When we compare paging to DASD vs. paging to Flash, paging I/O wait time is much lower with Flash Express, resulting in a significant reduction in the total elapsed time to capture a standalone dump.

Java Performance Benefits

The next set of tests was used to compare Java performance. Java's use of large pages is not new, so what is different and unique with Flash Express? Previously, Java supported the use of fixed large pages which remain backed by real storage the entire time they are allocated. (Fixed pages are always backed by real memory, whereas pageable pages are not backed by real memory until they are referenced. Pageable pages can also be paged to auxiliary storage if necessary.) In addition, only *authorized* users are allowed to use fixed large pages, and as a result, additional security controls and administration are required. No special authorization is needed to use pageable large pages.

With Flash Express running z/OS 1.13 on an IBM zEC12 along with the *RSM Enablement Web Deliverable*, both Java7 SR3 workloads and Java 6.0.1 SR4 workloads (and exploiters) can use 1MB pageable large pages to gain a performance benefit similar to that of fixed large pages.

- ***The use of pageable large pages (PLP) delivers a compelling performance benefit without requiring fixed real memory and without requiring a special security configuration.***

In addition, with Java7 SR3 and Java 6.0.1 SR4, users running Java 64-bit applications can use pageable large pages for their *JIT Code Cache.*

- ***WAS Day Trader 2.0 64-bit Java can gain an additional performance boost of up to 8% when using Pageable large pages***

In a Java 31bit Multi threaded workload test, applications using pageable large pages received a performance boost of 4%; 64- bit Java applications using the Day Trader benchmark experienced a significant boost of 8% from both the Java heap and JIT code cache. Other exploiters of Java, such as WebSphere and other middleware; will be eligible to leverage this performance boost with adoption of Java7 SR3. A summary of results follows:

Java Performance Test

Java 7 SR3	JIT	Java Heap	Multi Threaded	WAS Day Trader 2.0
31 bit	yes	yes	4%	
64 bit	yes		1%	3%
64 bit		yes	4%	5%

Figure 6. Java Benchmark results

DB2 Initial Performance Benefits

Our final set of tests with DB2 11 for z/OS showed that backing DB2 buffer pools and buffer pool control blocks with pageable large pages can allow the CPU to access memory more efficiently, lowering the CPU cost of accessing DB2 buffer pools.

DB2 supports both pageable and non-pageable buffer pools. IBM continues to recommend non-pageable buffer pools, especially for buffer pools that have high I/O rates, to avoid the CPU overhead associated with page fixing and freeing buffers during I/O. Flash Express lets you use large pages to improve the efficiency of memory access whether or not you use non-pageable buffer pools.

Note: We do not advocate expanding buffer pool sizes to the point where they cause paging. However, having buffer pools backed by large pages can result in more efficient access to memory.

The average customer should see about a 25% performance benefit moving from the IBM zEnterprise 196 to the zEnterprise EC12. We project that some customers using pageable buffer pools may see as much as a 3% additional performance improvement from using large pages for DB2 buffer pools and for buffer pool control blocks. Some DB2 customers may see as much as a 28% combined benefit moving their DB2 workload from z196 to zEC12 with Flash Express.

Conclusion

Consider use of Flash Express as a means of significantly improving the performance and the availability of your key systems that cannot afford availability disruptions. Flash Express is a very inexpensive way for organizations to meet demanding availability requirements. With faster page-ins from Flash, critical workloads may transition faster, allowing transactions to reach normal steady state quickly. SVC dump capture time can be dramatically reduced and the performance impact on other workloads (that would otherwise be non-dispatchable for longer periods of time) minimized. For companies whose delays can cost thousands of dollars per minute, this is a very compelling value proposition.

Java and DB2 performance can be enhanced as well using pageable large pages, which are only available on systems with Flash Express configured. In sum, Flash Express represents a way to improve availability and performance with a simple solution that is readily configurable, easy to use, and offering a fast return on investment.

IBM Systems and Technology Group
Route 100
Somers, New York 10589
U.S.A.
Produced in the United States of America,
5/2013

ZSS03073-USEN-02